
음성인식을 위한 성도 길이 정규화

지상문*

Vocal Tract Length Normalization for Speech Recognition

Sang-Mun Chi

요약

화자들 사이의 성도의 길이의 변이에 의하여 음성 인식기의 성능이 저하된다. 본 연구에서는 입력 음성에서 추출한 단구간 스펙트럼의 주파수축을 확대하거나 축소하여 음성인식기에 미치는 화자사이의 성도 길이의 영향을 최소화하는 방법을 사용한다. 성도의 길이를 정규화하기 위한 주파수 변환 함수로서, 선형의 주파수 변환 함수와 조각적 선형적인 변환 함수를 고려하였다. 또한, 커다란 성도길이의 변이에 따른 주파수축의 척도변화를 보다 효과적으로 모의할 수 있는 가변구간 조각적 선형함수를 제안한다. TIDIGITS 연결 숫자음 음성자료에 대하여 제안한 방법을 적용한 결과, 단어의 오인식률을 2.15%에서 0.53%로 크게 감소시켰으므로, 성도 길이 정규화가 화자 독립 음성인식기의 성능 향상에 필수적임을 알 수 있었다.

ABSTRACT

Speech recognition performance is degraded by the variation in vocal tract length among speakers. In this paper, we have used a vocal tract length normalization method wherein the frequency axis of the short-time spectrum associated with a speaker's speech is scaled to minimize the effects of speaker's vocal tract length on the speech recognition performance. In order to normalize vocal tract length, we tried several frequency warping functions such as linear and piece-wise linear function. Variable interval piece-wise linear warping function is proposed to effectively model the variation of frequency axis scale due to the large variation of vocal tract length. Experimental results on TIDIGITS connected digits showed the dramatic reduction of word error rates from 2.15% to 0.53% by the proposed vocal tract normalization.

키워드

Vocal tract length normalization, Frequency warping function, Speech recognition

1. 서론

고성능의 음성인식기를 제작하기 위해서는 음성신호로부터 언어정보만을 추출하고, 화자 종속적 변이와 음향환경의 영향을 제거하여야 한다.

화자 종속적 변이를 제거하기 위한 방법은 화자 적응과 화자 정규화로 나눌 수 있다. 화자 적응

은 인식 모델을 화자의 특성에 맞게 조정하는 방법으로 MAP (Maximum A Posterior)[1], MLLR (Maximum Likelihood Linear Regression)[2] 등이 연구되어 있으며, MLLR 방식이 비교적 적은 양의 적응 데이터로 우수한 성능을 보이는 것으로 보고되고 있다.

화자 정규화는 입력되는 음성 신호로부터 특징

*경성대학교 컴퓨터과학과

접수일자 : 2003. 7. 25.

벡터를 추출하는 과정에서 화자의 변이를 정규화하는 방법이다. 화자 사이의 변이의 원인은 억양, 톱바드 효과, 감정 등의 외부적인 요인과 화자의 성도의 길이와 모양의 차이에 기인하는 내재적인 요인으로 나눌 수 있으나[3], 주로 화자의 성도 길이에 따른 특징벡터의 변이를 최소화하는 VTLN (Vocal Tract Length Normalization)이 널리 사용된다. 화자적응방법이 화자정규화방법보다 효과적이지만, 각기 다른 측면에서 화자 종속적인 변이를 제거를 시도하므로 두 방법을 중복 적용하여 성능 향상을 얻을 수 있다 [3].

성도길이 정규화를 수행하는 처리 영역에 따라, 주파수 영역에서 직접 주파수 척도를 변화하는 방법[3, 4, 5], 캡스트럼 영역에서 선형변환을 사용하는 방법[5, 6, 7], 본래의 캡스트럼 계수와 주파수 척도를 변환한 후 추출한 캡스트럼 사이의 변환행렬을 구하여 적용하는 방법[5]이 있다. 비교실험[5]에서 보듯이 주파수 영역에서 직접 변환을 수행하는 것이 다소 좋은 성능을 보였다.

주파수축의 척도를 변환하는 비율인 워핑계수를 찾는 방법에는, 인식모델에 대하여 모든 가능한 워핑계수를 적용하여 추출한 특징벡터들을 강제 정합(forced alignment)하여 가장 높은 확률을 갖는 워핑계수를 찾아내는 방식 [4, 5]이 널리 사용된다. 인식모델 대신에 가우시안 분포함수의 조합이나 벡터 양자화를 통한 코드북을 사용하여[6, 8] 최적 워핑계수를 찾는 방법은 계산량을 감소시킬 수 있는 장점이 있으나, 실험[4]에서 보듯이 음성인식에 중요한 정보인 음성신호에 포함된 시간적인 정보를 사용할 수 없으므로 성능이 약간 저하되는 단점이 있다. 이밖에도 화자의 대표적인 음향학적 특징인 포먼트 (formant)들로부터 워핑함수를 찾아내는 방식[3, 9]이 있으나, 입력음성의 언어적인 내용에 포먼트 등의 음향학적 특성이 종속적이므로 견고한 추정이 어렵다는 단점이 있다.

본 논문에서는 성도길이 정규화를 수행하는 영역을 상대적으로 성능이 우수한 주파수 영역을 선택하여 직접 주파수 단위를 변화하는 방법을 사용하고 [3, 4, 5], 워핑계수를 찾는 방법에는 음성인식에 중요한 정보인 음성신호에 포함된 시간적인 정보를 이용할 수 있도록 인식 모델을 사용하는

방법 [4, 5]를 사용한다. 주파수축의 척도를 변환하는 함수는 일반적으로 사용되는 선형[4], 조각적 선형 (piece-wise linear)[5, 7] 함수와 함께, 음향학적 특징으로부터 주파수 변환 함수를 찾는 방식[3]에서 시도된 비선형 함수를 추정하여야 할 파라미터를 감소시켜서 가변구간 조각적 선형함수로 모의하고, 이를 인식모델을 사용한 통계적인 방법으로 추정한다.

II장에서 여러 가지 주파수 변환 함수에 대하여 기술하고, III장에서 성도 길이 정규화를 음성인식에 적용하는 절차에 대해 살펴본다. 성도 길이 정규화를 적용한 음성 인식 실험 결과를 IV장에서 고찰하고, V장에서 결론을 맺는다.

II. 주파수 변환 함수의 형태

화자의 성도 길이에 따라 음성의 스펙트럼은 주파수축으로 확대 또는 축소되어 나타난다. 예를 들어, 음성인식의 중요한 정보인 포먼트 중심주파수의 위치는 성도길이에 반비례하므로, 성도의 길이가 성인에 비해 짧은 어린이의 음성은 성인 음성을 주파수 축으로 확대한 결과로 나타난다. 그림 1은 이러한 예를 나타내었다. 그러나, 표 1에 나타내었듯이 그림 1의 스펙트럼 피크의 위치는 선형적으로 비례하지는 않고, 고주파영역에서 확대비율이 증가한다.

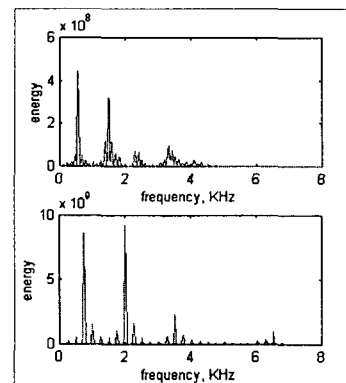


그림 1. 성인 남성(상)과 남자 어린이(하)가 발성한 "seven"내의 첫 번째 e의 스펙트럼
Fig. 1 spectrum of the first 'e' in "seven" uttered by adult male (top) and boy (bottom)

표 1. 그림 1에서의 스펙트럼 피크의 위치 (Hz)
Table. 1 The positions of spectral peaks in Fig. 1 (Hz)

피크위치	첫번째	두번째	세번째
성인남성	563	1489	2438
남자어린이	750	2000	3531
남자어린이 % 성인남성	1.33	1.34	1.45

이러한 비선형적인 확대비율의 예로서, 성인과 어린이의 음성에 나타난 첫 번째 세 개의 포먼트 주파수 F1, F2, F3에 관한 자료 [10]를 기초로 하여, 성인에 대한 어린이의 포먼트 주파수의 확대 비율을 모든 음소에 대해서 평균하여 표 2에 나타내었다. 표 2에서 보듯이 F1, F2의 확대비율은 유사하나, F3에서 더 크게 확대되었다. 물론, 이러한 비율은 음소의 종류에 따라 차이가 존재하지만, 전반적으로 고주파영역에서 확대의 비율이 증가하였다.

표 2. 여러 음소에서의 성인에 대한 어린이의 포먼트 주파수의 비율

Table. 2 The ratios of formant frequencies between children and adults for several phonemes

	F1	F2	F3
포먼트 평균비율	1.24	1.24	1.26
성인 포먼트 평균위치	539	1557	2585

VTLN에서는 이러한 성도 길이에 따른 스펙트럼의 확대와 축소를 정규화하기 위해, 주파수축의 척도를 변환하는 워핑 함수(warping function)를 사용한다.

$$F: [0, \pi] \rightarrow [0, \pi] \quad (1)$$

여기서, 정의역과 치역은 0Hz부터 표본화 주기에 따라 얻을 수 있는 최대 주파수까지이며, 성도의 물리적 특성에 따라, 단조 증가(monotonic increasing)하고, 연속인 함수를 사용한다.

워핑 함수의 첫 번째 예는 주파수축을 선형으

로 변환하는 함수이다.

$$F(\omega) = \alpha\omega \quad (2)$$

선형으로 주파수축이 변환되면, 변환된 신호는 원래의 입력 신호와는 다른 대역폭을 갖게 되므로, 주파수 범위의 경계에서 주의하여야 한다. α 가 1보다 작은 경우에는 변환된 신호의 대역폭이 원신호의 대역폭보다 작고, α 가 1보다 큰 경우에는 원신호의 대역폭보다 크다. 이러한 문제점을 해결하기 위해서 다음의 조각적 선형(piece-wise linear) 함수를 사용하여 같은 대역폭을 갖는 변환 함수를 사용하기도 한다. 본 연구에서는 α 가 1보다 큰 경우에 경계를 넘어서는 스펙트럼은 0으로 가정하였고, 넓은 주파수 범위를 가지는 필터 뱅크가 특징추출 과정에 사용되므로, 이러한 문제를 다소 완화할 수 있었다.

조각적 선형함수는 주파수 원점부터 특정 주파수 ω_0 까지는 선형으로 변환하고, 이후의 주파수에서는 기울기를 바꾸어서 본래의 스펙트럼과 변환된 스펙트럼의 최고 주파수를 일치시킨다.

$$F(\omega) = \begin{cases} \alpha\omega, & \omega \leq \omega_0 \\ \alpha\omega_0 + \frac{\pi - \alpha\omega_0}{\pi - \omega_0}(\omega - \omega_0), & \omega > \omega_0 \end{cases} \quad (3)$$

$$\text{여기서, } \omega_0 = \begin{cases} \frac{7}{8}\pi, & \alpha \leq 1 \\ \frac{7}{8\alpha}\pi, & \alpha > 1 \end{cases}$$

ω_0 는 음성인식의 중요정보를 대부분 포함할 수 있다고 가정할 수 있는 큰 값을 사용한다. $7\pi/8$ 일 경우는 8 kHz의 대역폭일 경우 7 kHz에 해당한다. 그림 2는 여러 α 에 대하여 조각적 선형 함수를 나타내었는데, x축과 y축의 단위는 8 kHz를 256개의 단위로 나누어 표시하였다.

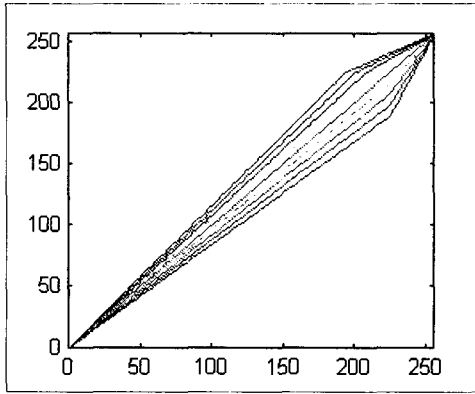


그림 5 : 여러 가지 다른 워핑계수에서의 조각적 선형 변환 함수들
 Fig 2: Piece-wise linear warping functions for several warping factors

표 1과 표 2에 보듯이 성도길이의 변이에 따른 주파수의 확대는 선형적이지 않으므로, 선형함수와 조각적 선형함수만으로는 성도길이의 변이를 제거할 수 없다. 본 연구에서는 주파수축의 확대비율이 크게 변하는 F2와 F3사이에서 주파수축의 확대비율을 바꿀 수 있는 가변구간 주파수 변환함수 $F(\omega)$ 를 제안한다.

$$\pi + \frac{\begin{matrix} a_1\omega, & \omega \leq \omega_1 \\ a_1\omega_1 + a_2 \cdot (\omega - \omega_1), & \omega_1 < \omega \leq \omega_2 \\ \pi - \frac{a_1\omega_1 - a_2 \cdot (\omega - \omega_1)}{\pi - \omega_2} (\omega - \pi), & \omega > \omega_2 \end{matrix}}{\pi - \omega_2} \quad (4)$$

여기서, $\omega_1 = \begin{pmatrix} \frac{f_1}{8} \pi, & a_1 \leq 1 \\ \frac{f_1}{8a_1} \pi, & a_1 > 1 \end{pmatrix}$ 이고, ω_2 의

경우에는 $a_1\omega_1 + a_2 \cdot (7\pi/8 - \omega_1) \leq 7\pi/8$ 일 때는 $7\pi/8$ 이고, 이외에는 $\frac{(7\pi/8 - a_1\omega_1)}{a_2} + \omega_1$ 로서, ω_2 에서의 변환함수의 값이 $7\pi/8$ 이상을 갖지 않도록 하였다.

주파수축이 확대되는 경우에는 고주파로 진행할수록 확대비율이 증가하는 특성을 반영하기 위하여, $a_1 > 1$ 일 경우는 $a_2 > a_1$ 가 되도록 기울기를

제한하였고, 같은 원리로 주파수축이 축소되는 $a_1 < 1$ 의 경우에는 $a_2 < a_1$ 가 되도록 제한하였다. 그림 3에 f_1 이 2 kHz이고, 여러 가지 a_1, a_2 에 대해 변환함수를 나타내었다. 그림 2와 마찬가지로 x축과 y축의 단위는 8 kHz를 256개의 단위로 나누어 표시하였다.

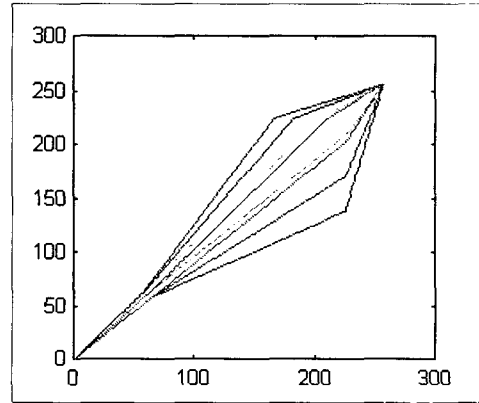


그림 6 여러 가지 워핑계수에서의 가변구간 선형 변환 함수들
 Fig 2: Variable interval piece-wise linear warping functions for several warping factors

III. 성도 길이 정규화

성도 길이 정규화를 수행하기 위해서는 입력 음성의 주파수 척도를 최적으로 변환하는 주파수 변환함수의 형태를 찾아야 한다. 최대 우도 (maximum likelihood)를 갖는 주파수 변환함수를 최적으로 가정한다.

$$\hat{F} = \arg \max_F \Pr(O^F | \lambda, W). \quad (5)$$

여기서, O^F 는 입력음성을 주파수 변환함수 F 로 주파수축의 척도를 변환하여 얻은 스펙트럼으로부터 추출한 특징 벡터열이고, λ 는 인식모델인 HMM (Hidden Markov Model), W 는 입력음성의 문자 표기이다.

식 (5)에서 우도(likelihood)를 최대화하는 주파

수 변환함수를 찾기 위한 닫힌 형식(closed-form)의 해가 존재하지 않으므로, 가능한 주파수 변환 함수 중에서 최대우도를 갖는 함수를 탐색한다.

선형 및 조각적 선형함수의 경우에는 형태를 결정하는 파라미터는 α 뿐이므로, 최적 α 를 추정하기 위해서 $0.6 \leq \alpha \leq 1.4$ 의 범위내에서 0.02간격으로 탐색하였다. 본 연구에서는 성인과 어린이의 성도길이의 변이의 범위가 크므로, 일반적으로 사용되는 탐색범위보다 넓은 범위를 탐색하였다.

가변주간 조각적 선형함수의 경우에는 f_1 를 1kHz, 1.5kHz, 2kHz 각각에 대하여, α_1 을 구간 $[0.6, 1.4]$ 내에서 0.04간격으로, α_1 이 1보다 작을 경우에는 $[\alpha_1 - 0.04, \alpha_1]$ 내의 α_2 를, α_1 이 1보다 큰 경우에는 $[\alpha_1, \alpha_1 + 0.04]$ 내의 α_2 를, α_1 이 1인 경우에는 $[\alpha_1 - 0.04, \alpha_1 + 0.04]$ 내의 α_2 를 0.02간격으로 탐색하였다.

VTLN을 인식모델의 학습단계와 인식하는 단계에 각각 적용하여 성도가 정규화된 인식모델과 입력음성에서 성도 길이의 변이를 제거할 수 있다. 학습단계에서는 입력음성의 문자 표기를 학습문장을 통하여 알 수 있으므로 이를 이용하여, 입력음성에 해당하는 HMM열 λ 를 찾아낼 수 있으나, 인식단계에서 문자표기를 사전에 알 수 없으므로, 이를 음성 인식을 통하여 찾는다. 즉, 입력된 음성에 대하여 주파수 변환을 하지 않고, 다음식을 사용하여 문자 표기 \hat{W} 를 일차적으로 얻어낸다.

$$\hat{W} = \arg \max_W \Pr(W|O) \quad (6)$$

얻어진 문자표기 \hat{W} 를 이용하여 최적 주파수 변환함수를 찾고,

$$\hat{F} = \arg \max_F \Pr(O^F | \lambda, \hat{W}). \quad (7)$$

최종적으로, 변환함수 $\hat{F}(\cdot)$ 로 입력음성의 주

파수축 척도를 변환하여 얻어진 특징벡터열 O^F 을 사용하여 인식을 수행한다.

IV. 실험 및 결과

4.1 실험 환경

여러 가지 VTLN 기법을 적용하기 위한 음성 인식 실험에는 TIDIGITS 자료를 사용하였다[11]. 성도 길이의 변이가 큰 경우에도 적용할 수 있는 주파수 변환함수의 형태를 알아보기 위하여, 성인 음성으로 HMM을 학습시켰고, 인식시에는 어린이 음성을 사용한다. 남성 55명, 여성 57명이 발성한 음성을 HMM의 학습을 위하여 사용하였고, 학습에 사용되지 않은 남자 어린이 25명, 여자 어린이 25명이 발성한 음성을 평가에 사용하였다. 성인은 17세부터 70세 사이이고, 어린이는 6세부터 15세의 범위이다. 각 화자는 11개의 영어 숫자(zero, oh, one, two, three, four, five, six, seven, eight, nine)로 이루어진 숫자열을 발성하는데, 한 자리 숫자들은 두 번씩, 그리고 두 자리부터 일곱 자리까지 연결된 숫자는 55개를 발성하였다.

16개의 Gaussian mixture와 10개의 상태를 가지는 연속분포 HMM을 사용하여 음성을 모델링하였다. 분석프레임의 길이는 32ms이고 16ms씩 이동하였고, 19개의 바야크 스케일 필터뱅크로부터 이산 코사인 변환을 사용하여 0차를 포함하여 11차까지 캡스트럼을 구하였고, 연속된 5개의 프레임의 캡스트럼을 사용하여 얻은 회귀 계수를 동적 파라미터로 사용하였다.

4.2 실험 결과

실험의 결과는 문장 오인식률과 단어 오인식률로 나타내었다. 문장 오인식률은 숫자열에 포함된 숫자들이 하나라도 오인식된 경우에 오류로 간주되고, 단어 오인식률은 다음과 같다.

$$WE = 100 \cdot \frac{\text{Sub} + \text{Del} + \text{Ins}}{\text{Total Number of Words}} \quad (8)$$

여기서, "Sub"은 인식시에 대체된 단어의 수, "Del"은 삭제된 단어의 수, "Ins"는 삽입된 단어의 수이다.

각 실험에서 Baseline은 VTLN을 적용하지 않은 기본 인식률이고, LIN은 선형함수, PL은 조각적 선형함수, VIPL은 가변구간 조각적 선형함수를 적용하여 VTLN을 수행한 결과이다. 인식대상의 화자마다 처음 발성한 3개의 문장으로부터, 주파수 변환함수의 파라미터를 추정하였다.

표 3에서 보듯이 주파수 변환함수를 적용한 것이 인식 성능을 크게 향상시켰고, 상대적으로 VIPL이 PL이나 LIN보다 우수하였다. 이는 주파수 척도의 비선형적인 변이를 더욱 효과적으로 정규화하기 때문이라 판단된다.

표 3. 여러 주파수 변환 함수의 음성인식 오류율 (%)
Table. 3 Speech recognition errors (%) for various frequency warping functions

방법 \ 오인식률	문장	단어
Baseline	8.50	2.15
LIN	3.46	0.63
PL	3.38	0.62
VIPL	3.17	0.53

표 3에서는 학습자료 전체가 성인의 성도특징을 대표한다고 가정하고, 학습자료를 주파수 변환하지 않았지만, 학습자료에 포함된 화자사이에 성도 길이의 변이가 존재하므로, VTLN으로 정규화하여 학습모델을 재차 학습할 수 있다. 표 4에서는 학습자료에 속한 각 화자가 발성한 10개의 문장으로부터 최적 주파수 변환함수를 추정한 후에, 이를 이용하여 각 화자의 음성을 변환하여 학습에 사용한 경우이다.

표 4. 학습자료에 VTLN을 적용한 후의 여러 주파수 변환 함수의 음성인식 오류율 (%)
Table. 4 Speech recognition errors (%) for various frequency warping functions after applying VTLN to training speech database

방법 \ 오류율	문장	단어
LIN	3.46	0.70
PL	3.41	0.67
VIPL	3.25	0.63

표 4의 결과는 학습자료를 변환하지 않은 결과인 표 3보다 성능이 조금 떨어지는데, 이는 식 (6)에 의해 추정되는 \hat{W} 의 부정확함에 기인한다. \hat{W} 의 추정에는 VTLN을 적용하지 않은 음성을 대상으로 수행되므로, 학습자료인 성인의 음성을 정규화할 경우에는 인식자료인 어린이 음성과 음향학적인 차이가 더욱 커지기 때문이다. 예를 들어, 학습자료에 VTLN을 적용하지 않고 어린이 음성을 인식할 때의 단어 오인식률은 표 3의 2.15%이지만, PL을 학습자료에 적용하여 모델을 학습시킨 후에 인식할 경우에는 3.4%로 증가하였다. 그럼에도 불구하고, 최종적으로 인식자료에 VTLN을 적용할 경우에는 VIPL이 LIN과 PL보다 성능이 우수하였다.

V. 결론

본 논문에서는 성도 길이의 변이에 따른 음성인식기의 성능저하를 완화하기 위하여 여러 가지 주파수 변환함수를 비교 실험하였다. 성도의 길이의 변이가 큰 경우에는 주파수 대역에서의 주파수 축의 확대 비율이 비선형적이므로 이를 정규화하기 위해 가변구간 조각적 선형함수를 제안하였다. 가변구간 조각적 선형함수는 주파수축의 확대비율과 확대비율을 변경위치를 화자의 특성에 종속적으로 최적화 할 수 있으므로 보다 효과적으로 성도길이를 정규화할 수 있다.

제안한 방법의 효용성을 검증하기 위하여 연속 분포 HMM을 사용하여 음성인식을 수행한 결과 2.15%의 단어 오류율을 0.53%로 크게 감소시켰다. 주파수 변환 함수의 모양에 따라서는 성능이 크게 차이가 나지는 않았으나, 제안한 방법이 보다 우수한 성능을 보였다.

제안한 주파수 변환함수는 단순히 성도의 길이가 변하는 음향환경뿐만 아니라, 감정이나, 롬바드효과에 의해 동일한 화자가 발성하더라도 포먼트의 이동이 비선형적으로 발생하는 음성에 대해서 효과적인 적용이 가능할 것으로 예측되며, 추후의 연구에서는 이런 음성자료에 대한 실험을 수행할 예정이다.

참고 문헌

- [1] C. H. Lee, C. H. Lin, and B. H. Juang, "A study on speaker adaptation of continuous density HMM parameters," Proc. ICASSP, 1, 145-148, 1991.
- [2] C. Leggetter and P. Woodland, "Maximum likelihood linear regression for speaker adaptation of continuous density hidden markov models," Computer Speech and Language, 9, 171-185, 1995.
- [3] E. B. Gouvea, "Acoustic-Feature-based frequency warping for speaker normalization," Thesis, Carnegie Mellon University, 1998.
- [4] L. Lee and R. C. Rose, "A frequency warping approach to speaker normalization," IEEE Trans. on Speech and Audio Processing, 6 (1), 49-60, 1998.
- [5] L. F. Uebel and P. C. Woodland, "An investigation into vocal tract length normalization," Proc. EuroSpeech, Vol. 6, 2527-2530, 1999.
- [6] J. McDonough, W. Byrne, and X. Luo, "Speaker normalization with all-pass transforms," Proc. ICSLP, paper no 869, 1998.
- [7] M. Pitz, S. Molau, R. Schluter, and H. Ney, "Vocal tract normalization equals linear transformation in cepstral space," Proc. EuroSpeech, E31, 2653-2656, 2001.
- [8] 신옥근, "DHMM 음성 인식 시스템을 위한 양자화 기반의 화자 정규화," 한국음향학회지, 제 22권 제 4호, 299-307, 2003.
- [9] Y. Ono, H. Wakita and Y. Zhao, "Speaker normalization using constrained spectral shifts in auditory filter domain," EuroSpeech, 1, 355-358, 1993.
- [10] T. D. Rossing, The science of sound, Addison-Wesley publishing company, p. 320, 1989.
- [11] R. G. Reonard, "A database for speaker-independent digit recognition," Proc. ICASSP, 3, 42.11/1-4, 1984.

저자 소개



지상문(Sang-Mun Chi)

1991년 : 서울대학교 수학교육과 (학사)

1993년 : 과학기술원 수학과 (석사)

1998년 : 과학기술원 전산학과 (박사)

1993년 ~ 2000년 : 삼성전자 (선임연구원)

2000년 ~ 2001년 : L&H (책임연구원)

2001 ~ 현재 : 경성대학교 컴퓨터과학과 조교수

※관심분야 : 음성처리, 패턴인식